DOCUMENT RESUME

ED 186 472 TM 800 177

AUTHOR Hartman, Allen S.
TITLE Overview of a Policy Evaluation Design for Determining the Effectiveness of a State Minimum Competency Policy.
PUB DATE Apr 80
NOTE 10p.; Paper presented at the Annual Meeting of the American Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Basic Skills; Educational Assessment; *Educational Policy; Educational Testing; Elementary Secondary Education; *Information Needs; *Minimum Competency Testing; Politics; *Program Effectiveness; Program Evaluation; Public Policy; Research Problems; *School Responsibility; *State Programs
IDENTIFIERS Massachusetts

ABSTRACT

A program is described for the evaluation of the basic skills policy of the State Board of Education in Massachusetts. The 1978 policy of the Board is to take effect in 1980-81 and will require annual assessment of student skills at three grade levels in all schools in five curriculum areas, but leaves the setting of standards to local option. Various groups have expressed strong positions for and against this policy and have raised a variety of questions about possible intended and unintended effects. Questions about the impact of this policy are grouped under four major headings: (1) what is the impact on student basic skills achievement; (2) what other impact on students does the policy have; (3) what is its impact on school practices; and (4) what is the impact on school/community support for the policy? Questions are raised about what data to collect, how to collect it, and what to do with it. Limits on rational evaluation strategies are discussed, but the usefulness of evaluation for decision making stressed. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186472

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Allan S. Hartman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

OVERVIEW OF A POLICY EVALUATION DESIGN FOR DETERMINING THE EFFECTIVENESS OF A STATE MINIMUM COMPETENCY POLICY

Presented at the Annual Meeting

of the

American Educational Research Association

April, 1980

Allan S. Hartman
Bureau of Research and Assessment
Department of Education
31 St. James Avenue
Boston, Massachusetts 02116

TM800 127

Background - Purpose

Over 36 states are currently operating under some form of minimum competency policy. These policies, which typically emanate either from legislative or state board initiative, vary, of course, in how much they do or do not require of state education departments and local education agencies. They vary in other respects also. Some are extremely specific about what competencies students must master; some leave it to the school system. And the specifics of testing, tests, standards, and what is to happen with those who do not meet standards varies similarly.

Yet, despite the variation (or, perhaps because of it), there is one persistent concern common to all states engaged in minimum competency activities. That is...is the policy effective? Does it achieve its intended purpose while minimizing unintended negative side effects or outcomes? In short, over the years, it may reasonably be expected that states will be called to account for the impact of minimum competency policies adopted in the seventies. And to do this properly states will have to include among their plans some explicit evaluation design for answering the questions that may be expected to be asked.

This paper will provide an overview of the beginning one state is making in preparing to answer these questions. This paper will first describe Massachusetts' competency policy following which it will discuss the strategy for finding out what the major questions are. This will lead to a discussion of some of the major questions and what these questions suggest for building a data base and an analytical framework. The paper will conclude with some observation on the limits of rational evaluation strategies.

Massachusetts' Minimum Competency Policy

Massachusetts' response to the minimum competency debate of the seventies was in the form of an explicit policy statement passed by the State Board of Education in late 1978. For the time being Massachusetts decided not to require the passing of a test as a prerequisite for high school graduation. Instead, the Board opted for a Policy whose purpose was stated as assisting:

> "...all students in achieving mastery of basic skills prior to high school graduation through the provision of appropriate curriculum, instruction, and evaluation."

Diagnosis and follow-up of learning problems was emphasized and a high amount of local involvement, discretion and responsiblity called for.

Specifically, Massachusetts requires that student achievement of minimum standards in reading, writing, mathematics, listening and speaking be assessed locally at least once at each of three levels: early elementary (K-3), later elementary (4-6), and secondary (7-12). The State leaves to

local schools the responsibility for defining both objectives and standards at the elementary levels but requires, at the secondary level, that schools incorporate state specified objectives in reading, writing, mathematics, listening and speaking into their programs. Schools, however, have the resporsibility for setting standards (cut-off points, levels of attainment, etc.) at each of the three levels noted above.

In addition to establishing objectives and minimum standards at all levels in all basic skills areas, schools have the responsibility for determining specific grades to be tested and the testing schedule, administering the tests, revising regular instruction to provide appropriate follow-up with students not meeting local standards, evaluating the impact of their programs, and reporting annually, the results to the State and the public at each of the three grade levels. Implementation of the testing programs begins in 1980 for reading, writing and mathematics and 1981 for listening and speaking.

At the elementary levels the choice of assessment instruments is the responsibility of the local school districts. At the secondary level instruments may be (1) tests developed by the state; (2) state approved commercial instruments; or (3) tests developed by the local districts that are comparable to either (1) or (2). The state has provided schools with a list of approved commercial tests and criteria by which a local district may evaluate its own tests. In addition, state tests and equivalent forms have been developed which may be used by schools selecting the first option noted above.

As mentioned earlier, schools must report, annually, on the results of their testing at three grade levels. Specifically, schools must report "the number and percentage of students overall and by race, sex and linguistic minority who have and have not achieved the minimum standards for each basic skill." Reports are to be filed with the Department and made public.

Finally, school systems must submit to the Department of Education, beginning this year, their plans for doing the preceding. At the elementary level only an assurance is required but at the secondary level complete program plans must be described. This description must include information on public participation, minimum standards, evaluation instruments, follow-up activities and school district plans for monitoring, evaluating and modifying their program plans. The Commissioner of Education, following recommendations made by regional advisory groups, will either approve or disapprove local school plans.

While there are other aspects of the Policy concerning exemptions, the review process and such, it should be evident to the reader that Massachusetts' policy places a high premium on local autonomy and decision-making and a low state profile. The state's role is primarily technical support and the overseeing and monitoring of local district implementation of competency programs. Yet even in this context, one where there promises to be a wide variety of practices, the question remains, will the total of such activity lead to improved student competency? Will the policy have the intended effect of aiding students in achieving mastery of certin basic skills before graduating from high school? Even if it can be shown that the Policy does result in improved student competency, how will decision-makers in a few years be able to consider whether policy changes are warranted or not? In short, how will Massachusetts be able to evaluate the effectiveness of the Policy as a basis for well-informed policy reconsideration?

Locating the Policy Questions

The very fact that the formulation of the Massachusetts Policy on Basic Skills Improvememt represented a series of compromises among various groups and interests within the state, suggests that with or without formal evaluative information, some of these groups of interest will press their claims for Policy changes into the State Board of Education or the legislature within a few years of the Policy's implementation. In fact, the State Board of Education, recognizing the strong stance of some educators and lay persons towards a mandatory graduation requirement, expiicitly indicated in the Policy its intention to reconsider the issue no later than June, 1983.

But there were other issues also, with fluent proponents, which have been temporarily laid aside. Elementary educators feel that the policy does not concentrate sufficiently (at all, some would contend) on the early years of a child's development of basic skills. Some special educators feel that the state's neutral stance towards including or excluding special needs students may deny many of these students appropriate instructional services. Some educators and lay persons feel that the assessment of listening and speaking skills will escalate school testing costs. Some concerned teachers and curriculum personnel feel that the stress on basic skills will contract the regular curriculum. A number of educators feel that in some communities' tests results will be used to evaluate teachers. And so on and so on!

It became apparent over the past year, as schools began making plans to implement the Policy (which calls for district implementation beginning in 1980/81), that systematic data collection focused on critical policy questions had to begin immediately. At the very least it would be important to identify baseline information so that changes could be recorded over the years. As important, it became clear that many policy questions dealt not only with outcomes but also with side-effects, processes and contexts. This type of information, fugitive as it might be, had to be sought as conscientiously as the more easily available performance data.

Another issue that had to be confronted was that of identifying the future decision-makers...those whose thoughts, ideas and recommendations would bear most significantly on future policy changes. With the exception of the legislature (whose intervention is not anticipated), the most important decision-makers are the State Board and Commissioner of Education. While as individuals and as a group they are influenced by many outside groups, the two most influential groups that will have their ear are the regional education councils and a statewide Advisory Committee on Basic Skills Improvement. The former, composed of both lay persons and educators, were established by the Board to advise on regional needs and are built into the Policy as the key group to both recommend approval or disapproval of local school plans and "...for reviewing the implementation of the Policy on Basic Skills Improvement in each region."

The latter group, the Advisory Committee, advised the Board initially during the period of policy formulation. With the establishment of the Policy that Advisory Committee was designated as the "...principal advisory committee (to the Board) on matters related to the Policy." The Committee is composed of 40 members representing major educational constituencies such as administrators, parents, teachers, students, school committees and specialized interests such as bilingual, educators, special education, guidance and basic skills subject areas.

Since both groups represent a diversity of interests and since both regularly report to the Board on their concerns and recommendations, it seemed, at least as a starting point, reasonable to screen out the major policy questions through an iterative process of polling and querying these groups as to the major questions they would want answered over time. Similarly, the Commissioner's concerns and the types of information he would seek had to be elicited. It is believed that most of the questions of concern to other interest groups will be identified through continuous back-and-forth shaping of an evaluation design with the councils and the Advisory Committee.

Policy Questions and Data

In 1979, with advice and suggestions from the Advisory Committee, some benchmark information was gathered through a statewide student sample, on basic skills achievement in reading, writing and mathematics. Performance data was analyzed by subject area and by such selected reporting group categories as sex, race, kind of community, linguistic background, socio-economic status, program enrollment and responses to a variety of questions. In addition data was collected and analyzed so as to describe selected characteristics and perceptions of students about their schools and to describe similar information from the principals in the 55 sample schools.

While much of this information will be gathered again in a few years, much of it may not be gathered again unless the groups influencing the State Board or the Board itself indicate a desire for it. Some would contend that is is important to gather certain types of information whether or not anyone expresses a need for it. The position in this evaluation design is one of consumer orientation. What do decision-makers or those advising decision-makers want to know? And even if there is a responsibility of the evaluator to attempt to educate the client in what he should want, ultimately the decision rests with the client. The perception of the importance of information to the client is where the evaluator starts...and maybe ends!

With this in mind the Bureau of Research and Assessment set out to find out what information both the regional councils and the Advisory Committee wanted and what value they placed on this information. The reasons for determining the value placed on the information by these groups is that values indicate whether or not a piece of information is relevant to their decision system. The job of the Bureau (i.e. evaluator) is to supply knowledge or information relevant to decisions.

Specifically, the Bureau sent out a list of initial questions that seemd to emcompass all major information needs of decision-makers. The initial list was drawn from information collected dealing with the concerns of many persons and agencies about the Policy and the possible consequences resulting from its implementation. This list was supplemented with information gathered from the literature which focused critically on the possible negative and positive outcomes of minimal competency mandates.

Whether the Policy has achieved its major purposes (as noted earlier) and whether and in what ways it should be changed will be determined, in part, by how decision-makers react to information about the impact of the Policy. And, the primary impact of the Policy, about which information will be needed, will be on students, schools and community. The questions related to the impact of the Policy on these groups may be outlined as follows:

I. What has been the impact on student basic skills achievement?

- on levels and changes in achievement?
- on various categories (e.g. minorities) of students, at various grade levels and among various basic skills areas?
- on students with different school histories (programs, courses, instructional practices, etc.)?

II. What other impacts on students has the Policy had?

- on student learning, experiences, and achievement in other areas?
- on student perception and attitudes about the school (options, challenges, etc.) and teachers (effectiveness, knowledge, etc.)?

III. What has been the impact on school practices?

- on instructional offerings and materials?
- on special programs (special education, Title I, bilingual, etc.)?
- on perceptions and attitudes about school climate (i.e. learning orientation, goal direction, etc.)?
- on attendance, dropouts, promotion and graduation?
- on school organization and administrative practices (grading, tracking, scheduling, supervision, course offerings, etc.)?
- on test usage, student assignment, and amount of stress on basic skills?
- on quantity/quality of personnel available, personnel assignment, in-service training and basic skills "time on task"?

IV. What has been the impact on school/community support for the Policy?

- on costs and management burden?
- on school/community perception of improved school learning?
- on attitudes concerning maintainence of programs?

The final listing of the types of information to be collected will depend on the weight the various decision-makers place on the above. This will be determined as the Bureau interacts, in an iterative manner, with these groups following the initial survey. With each interaction the information to be collected will be listed with increasing specificity. For example, the information to be collected on the basic skills achievement of various categories of students would be displayed as follows:

I. Information to be collected on the basic skills achievement of various categories of students:

A - Minorities

- Racial
- Linguistic

B - Socio-Economic/Geographic Status

- Low, Middle, High
- Urban, Suburban, Rural

C - Program Types

- College, Vocational, Business, General, etc.
- Special Education
- Title I

Data Collection, Analysis and Reporting

The decision as to what data to collect is, obviously, only the starting point in an evaluation design. How data will be gathered and what will be done with it are the other major components.

The initial departure point for data collection will require matching the data to be collected with specific data instruments administered at specific times. Currently, there are two major data collection instruments in place to gather impact data:

1) Annual Reports from school districts. These reports, which will be submitted to the Department of Education, will provide yearly information on student achievement (including breakdowns by various student categories), and district instructional activities.

2) Periodic state tests and surveys. These instruments will assess student basic skills achievement statewide (using a statewide probability sample) and gather data on student and personnel perceptions and attitudes and collect school organization and program information.

Whether other data collection instruments will have to be designed and administered will become clear following the design of the above instruments. Presently, the Division of Curriculum and Instruction is designing an instrument to monitor school system implementation of programs. It is possible, therefore, that additional data will be provided through this form of data collection.

Separate analyses will be performed within each instrument and subsequently between instruments. While is can be expected that a wide variety of analyses will be performed on the data resulting from the administration of each instrument, the specific identification of what analyses will be undertaken will depend on a final determination of what types of information are to be reported. It can be anticipated, however, that analyses will include:

- descriptive information (e.g. p-values, mean percentages, cross tabulations, etc.), statewide, by strata, and by student and school characteristics

- correlational information (e.g. between student characteristics and test performance)

- statistical inferences that may or may not be made about observed and reported data

Separate data collection, analysis and reporting will be carried out during 1980/81, 1981/82, and 1982/83. This process will permit some longitudinal comparisons as well as allowing for cross sectional analysis over a three year period.

Reports to decision-makers will be made annually. Reporting will include information on findings and analysis. To help in interpretation, the findings and analysis will be grouped under the various questions noted earlier.

Limits on rational evaluation strategies

The major limitations in carrying out the evaluation strategy described are technical, political and economic. How much these limitations will prevent this evaluation from serving its purpose is not yet clear. It is hoped that developing and sharing the proposed evaluation design with decision-makers early... before the Policy is actually implemented and the collection of data begins will help reduce misunderstandings and opposition.

Technically, the major limitations have to do with the appropriateness, sufficiency and quality of the data collected by a state agency. Much of the data will be of a self-report nature with all the inherent limitations such data presents. Much of the data will be reported on the basis of the administration of a variety of test instruments which present difficulties in equating and interpretation. Some information will be derived from proxy data; an attitudinal survey on student perceptions rather than the use of observational instruments. While these limitations are real they are probably not insuperable obstacles to overcome. They simply suggest cautions in making much of small differences in findings.

Technically, however, the major cautionary note has to do with not attributing causality to the findings. Analysis of data will produce many interesting and suggestive relationships among a variety of variables. Yet the design, while not of the ex-post facto kind, is certainly not experimental or even quasi-experimental. The design's limit is similar to that of a time-series design without the benefit of any control groups. Movement and change can be recorded but causal inference in limited.

Politically, the evaluation is at the mercy of the varying viewpoints, vested interests and contests of power and influence that plague all evaluations that have the potential of significantly impacting on important decisions. The very fact that there was little if any demand for a formal evaluation of the Policy's impact...that the Bureau of Research and Assessment had to suggest and initiate preparations for such an evaluation on its own and without responding to the recommendation of decision-makers suggest a fairly low level of enthusiasm for such an undertaking.

And all phases of the evaluation appear to bring forth fears and concerns on the part of decision-makers and those whom they represent! Decisions on what data should be collected have elicited strong negative reactions to gathering student perceptual and attitude data about teachers. Analysis and reporting of performance data have brought forth concerns about the misuse of data in making invidious comparisons or bringing about the firing of teachers.

The reporting of minority data leads to concern over reinforcing stereotypes. And the use of tests erupts into vigorous accusations about test reliability, validity and bias. In short, various sorts of resistances to evaluation may lead to the eventual downgrading of evaluation results and their importance.

Economically, the evaluation is limited by the amount of money and personnel available to implement a comprehensive, quality design. As noted, the evaluation will rely heavily on school system reports and statewide tests and surveys conducted with student samples. Funds to supplement these instruments and procedures with case studies, interviews and systematic school and classroom observations are not available. Neither are sufficient funds available to verify the accuracy of data collected nor to conduct as wide an analysis as would be desirable. While most evaluations probably suffer from fiscal limitations, this can result in providing inadequate information and overly cautious interpretations to decision-makers.

Despite the limitations noted, the evaluation outlined can be useful to decision-makers in making the decisions they will have to make. At the very least, such an evaluation can help reduce wild or distorted claims about the Policy and its effects. More likely it can assist in refining judgements. If such evaluation can lead to an appreciation on the part of decision-makers as to the value of evaluation findings, and therefore, attenuate or eliminate political considerations it will have served its purpose.